Draft version 1.1

ATLAS NOTE

March 29, 2010

Using bin-by-bin corrections for track-based underlying event measurements in pp collisions at 900 GeV

A. Buckley[a], S. Chekanov[b], G.A. Hare[c], D. Kar[d], A. Moraes[e], J. Nielsen[c], J Proudfoot[b], R Yoshida[b], S. Wahrmund[d]

[a]*University of Edinburgh*
[b]*HEP Division, Argonne National Laboratory, 9700 S.Cass, Argonne,IL 60439,USA*
[c]*Santa Cruz Institute for Particle Physics*
[d]*TU Dresden - Institute of Nuclear and Particle Physics*
[e]*University of Glasgow*

Abstract

Measurements of charged particle densities using minimum-bias events from proton-proton collisions at a center of mass energy of 900 GeV are presented. The densities and the average transverse momenta of charged particles are shown as a function of the transverse momenta of a leading in p_T charged particle, concentrating on phase-space regions which are most sensitive to the underlying event. The data are unfolded using a bin-by-bin correction procedure and compared to several Monte Carlo (MC) predictions with different treatments of the underlying event.

1 Introduction

Measurements which have sensitivity to the soft underlying events are important for early physics in ATLAS as they allow us to understand data in terms of the Monte Carlo models to be used to extract potential new physics. The ATLAS collaboration has already published an article on charged multiplicities [1], concentrating on global event characteristics. We extend this measurement by studying charged particle densities in regions which are most sensitive to the physics of the underlying event. The measurement is performed in three regions of phase space, as shown in Fig. 1, where the "transverse" region is the region which is the most affected by the soft QCD processes responsible for the underlying events. Unlike the measurements presented in [1], the current measurement is done using a bin-by-bin correction as discussed below.

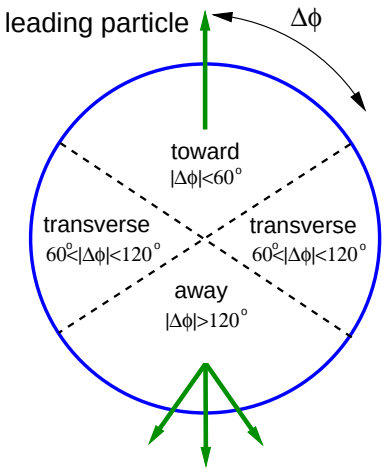

Figure 1: A schematic representation of the phase-space regions used for the studies presented in this note.

2 Data selection and MC samples

The 900 GeV data used in this analysis were collected during 2009 and the events to be analysed were selected using an identical procedure as described in [1]. As in [1], only primary tracks are selected with $p_T > 500$ GeV and $|\eta| < 2.5$[1]. The datasets used were those obtained from the first reprocessing.

The analysis was done using ESD/AOD files, converting them into the NtupleMaker [2] format to reduce the size and the increase the data reading rate. The NtupleMaker keeps information on tracks and truth particles in the form of a TLorentzVector-derived class (unlike D3PD which are based on vectors of numbers). After setting loose cuts on tracks and truth particles ($p_T > 0.45$ GeV), the size of the output files was reduced by a factor 8 compared to the standard D3PD and a single-pass over the data was reduced by a factor five. As a cross check, a similar analysis was performed by rerunning over D3PDs and converting its record into the NtupleMaker tree.

For the MC samples, the following sets are used:

[1]The ATLAS reference system is a cartesian right-handed coordinate system, with the nominal collision point at the origin. The anti-clockwise beam direction defines the positive *z*-axis, while the positive *x*-axis is defined as pointing from the collision point to the centre of the LHC ring and the positive *y*-axis points upwards. The azimuthal angle ϕ (radians) is measured around the beam axis, and the polar angle θ is the angle measured with respect to the *z*-axis. The pseudorapidity is defined as $\eta = -\ln\tan\theta/2$. $p_T >$ is the track momentum transverse to the beam direction.

mc09_900GeV.105001.pythia_minbias.merge.AOD.e500_s655_s657_d257_r1023_r1051
mc09_900GeV.105003.pythia_sdiff.merge.AOD.e466_s655_s657_d257_r1023_r1051
mc09_900GeV.105004.pythia_ddiff.merge.AOD.e466_s655_s657_d257_r1023_r1051

mc09_900GeV.108310.pythia_minbias_DW.merge.AOD.e504_s655_s657_d257_r1023_r1051
mc09_900GeV.108311.pythia_sdiff_DW.merge.AOD.e504_s655_s657_d257_r1023_r1051
mc09_900GeV.108312.pythia_ddiff_DW.merge.AOD.e504_s655_s657_d257_r1023_r1051

mc09_900GeV.108313.pythia_minbias_Perugia0.merge.AOD.e504_s655_s657_d257_r1023_r1051
mc09_900GeV.108314.pythia_sdiff_Perugia0.merge.AOD.e504_s655_s657_d257_r1023_r1051
mc09_900GeV.108315.pythia_ddiff_Perugia0.merge.AOD.e504_s655_s657_d257_r1023_r1051

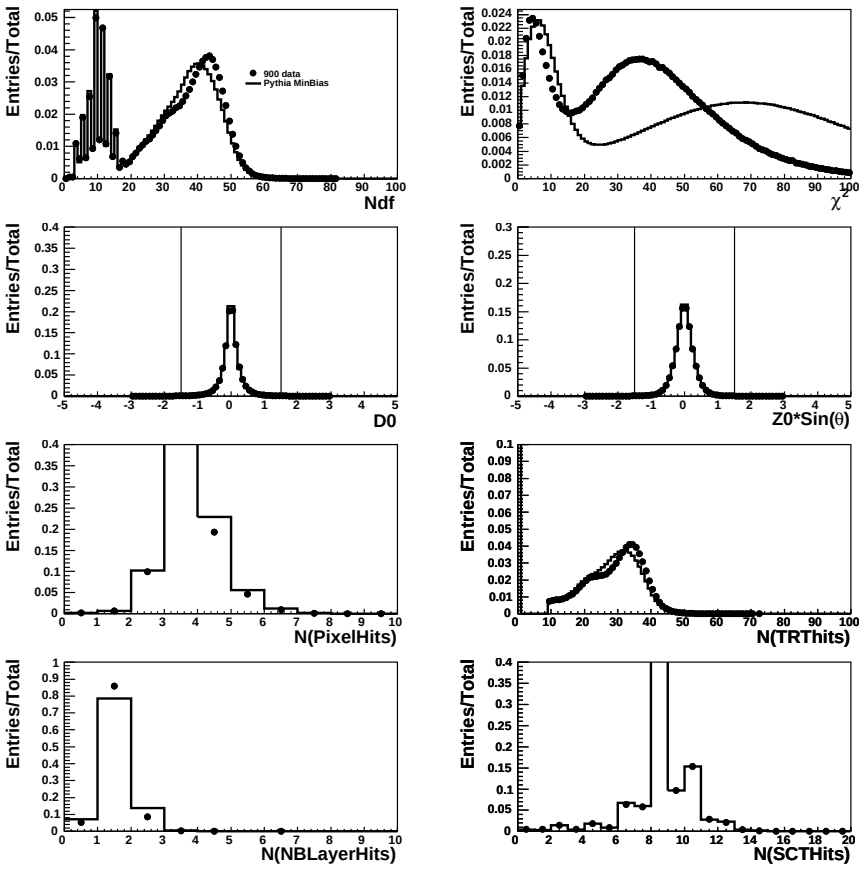

Figure 2: Several distributions which illustrate the track selection: the number of degrees of freedom on the track (Ndf), the track fit chisquared (χ^2), the track perigee to the event vertex (d0), the scaled track distance of closest approach in z along the beam line ($z0\sin(\theta)$), the number of hits on the track in each tracking detector (PixelHits, TRThits, BLayerHits, SCThits). The vertical lines show the cuts applied on the distance of the closest approach to select primary tracks.

The main analysis is done with the first MC set, while the other two tunes are used for systematics studies as discussed below. We do not use the PHOJET MC model [3] for the reasons to be described in Sect. 6.

To provide high statistics Monte Carlo samples for comparison with the unfolded data and in order to

study other MC tunes, the NtupleMakerTruth [4] was used to generate ROOT tree with TLorentzVector records. The official Monte Carlo production option files were used. The truth level to be shown on all figures was generated with the statistics a factor 5-10 larger than that used for data unfolding.

Figure 2 shows several track-quality in data distributions for the selected sample. Figure 3 shows shape comparisons for a few basic distributions between data and PYTHIA after the selection cuts. The standard ATLAS MC09 tune was used. There is a good agreement between data and Monte Carlo predictions for the shapes. Some small discrepancies are seen for small multiplicities. These are attributed to a contribution from diffractive events (will be discussed below).

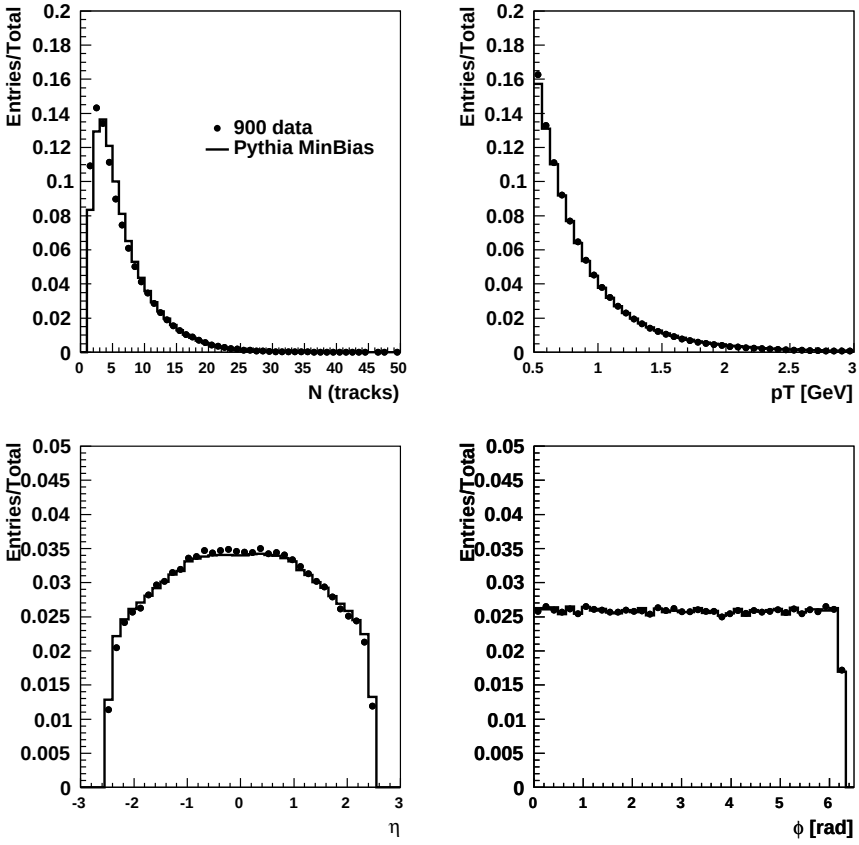

Figure 3: A shape comparison between data and PYTHIA after the detector simulation. All distributions are normalized to 1. A small discrepancy between data and MC in the region with small number of tracks is attributed to the presence of diffractive events (not included to the MC simulation).

One of the main motivations for this analysis is to reconstruct particle densities, rather than doing a shape-comparison analysis. A density is defined by dividing the number of entries in a given bin by the total number of events and by the bin size, where the total number of events is calculated as the number of events which have a leading tracks with $p_T > 1$ GeV. Figure 4 shows the shape distributions in Fig. 3 transformed into particle densities as defined above. Good agreement between data and Monte Carlo is still evident.

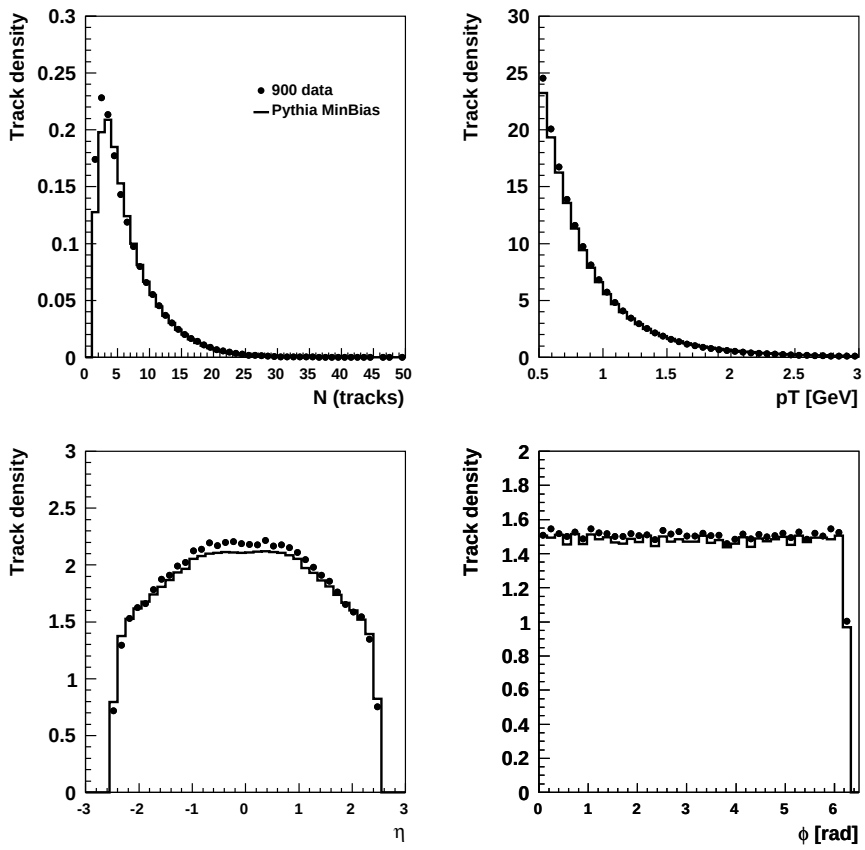

Figure 4: A comparison between data and Monte Carlo for track densities as described in the text.

3 Uncorrected distributions

In this section we will consider the observables we wish to study, but reconstructed at the detector level without unfolding to the truth level.

Figure 5 shows the density of tracks as a function of the azimuthal angle between the leading track and any other track in an event with the $p_T > 0.5$ GeV. The comparison to Monte Carlo predictions was performed with the PYTHIA model after full detector simulation. The PYTHIA model was tuned using the standard MC09 ATLAS tune and the Perugia0 tune. The density per unit pseudorapidity is defined as

$$\frac{N}{(\eta_{max} - \eta_{min})} \frac{1}{N_{ev} \delta\phi},$$

where N is the number of entries in the $\delta\phi$ bin of the size 0.16 rad and $\eta_{max} - \eta_{min} = 5$ represents the full pseudorapidity range and N_{ev} is the number of events triggered by a track with p_T above some value.

Although the general shape of the particle density distribution given by the Monte Carlo models agrees with the experimental data, we note that the data contain a higher track density at large angle to the lead track.

Figure 6 shows the number of tracks in an event in a given $p_T(lead.)$ bin per unit range in ϕ and η. This normalized density distribution is calculated as:

$$\frac{N}{(\eta_{max} - \eta_{min})} \frac{1}{N_{ev} \Delta\phi},$$

where N is the number of entries in bins of $p_T(lead.)$, N_{ev} is the number of events, and $\Delta\phi$ is the range in ϕ. In the case of the toward, away and transverse regions, $\Delta\phi = 0.33$ rad.[2]

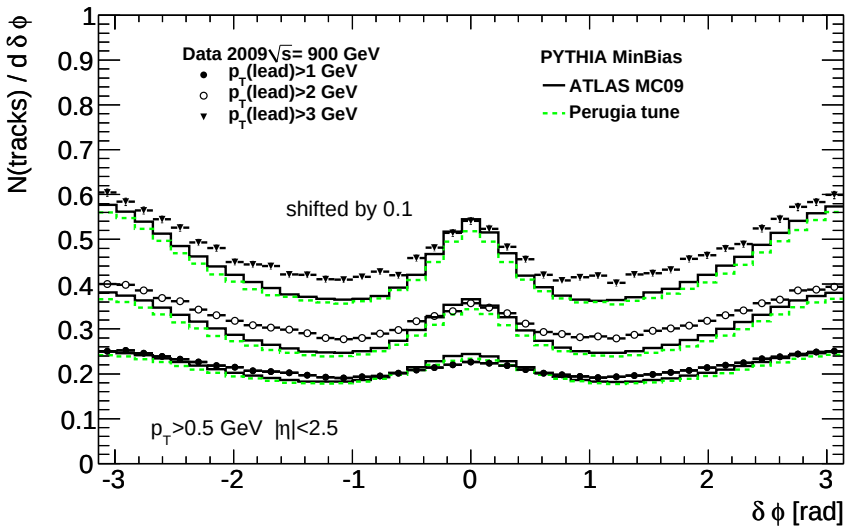

Figure 5: Track densities a function of the distance in the azimuthal angle between the leading track and any other track in an event. The density for $p_T(lead.) > 3$ GeV was shifted by 0.1 for a better separation from the other distributions.

[2] It should be mentioned that the above distribution is a normalized frequency distribution expressed in terms of the $p_T(lead.)$ variable as no division by the bin size is assumed.

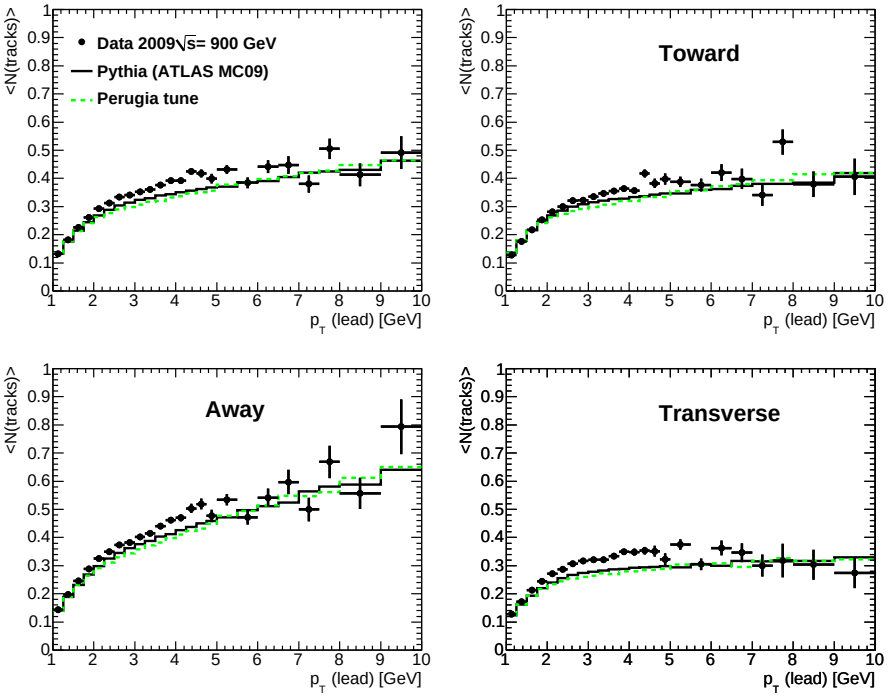

Figure 6: The average number of tracks per event per unit interval in η,ϕ as a function of $p_T(lead.)$ for the different regions defined in Fig. 1.

Figure 7 shows the average track transverse momentum in an event as a function of $p_T(lead.)$. Both Fig. 6 and 7 show similar features; the Monte Carlo prediction lies systematically below the data as $p_T(lead.)$ is increased. This is particularly true in the transverse region and reflects the corresponding difference in particle density between Monte Carlo and data seen in this region (Fig. 5).

4 Diffraction

To understand the contribution from diffractive events, we have used the PYTHIA MC09 models with single (SD) and double diffraction (DD) as described in Sect. 2. The minimum-bias events were mixed with the single and double diffractive events in accordance with the corresponding cross sections defined by PYTHIA.

To illustrate the contribution from diffractive events, Fig. 8 shows the visible cross section differential in p_T at the detector-level as a function of $p_T(lead.)$ from 0.5 to 2 GeV. The shaded histograms show the single and double diffractive contribution to the visible cross section in PYTHIA. The diffractive contribution is at the level of 13% (SD) and 2% (DD) for $p_T(lead.) > 0.5$ GeV. It decreases to 3% and 1%, respectively, for $p_T(lead.) > 1$ GeV. The diffractive contribution is negligible for $p_T(lead.) > 2$ GeV (below 1%). With this observation in mind, the analysis was done for $p_T(lead.) > 1$ GeV where the diffractive contribution is smaller than the overall systematic uncertainty (to be discussed in Sect. 6).

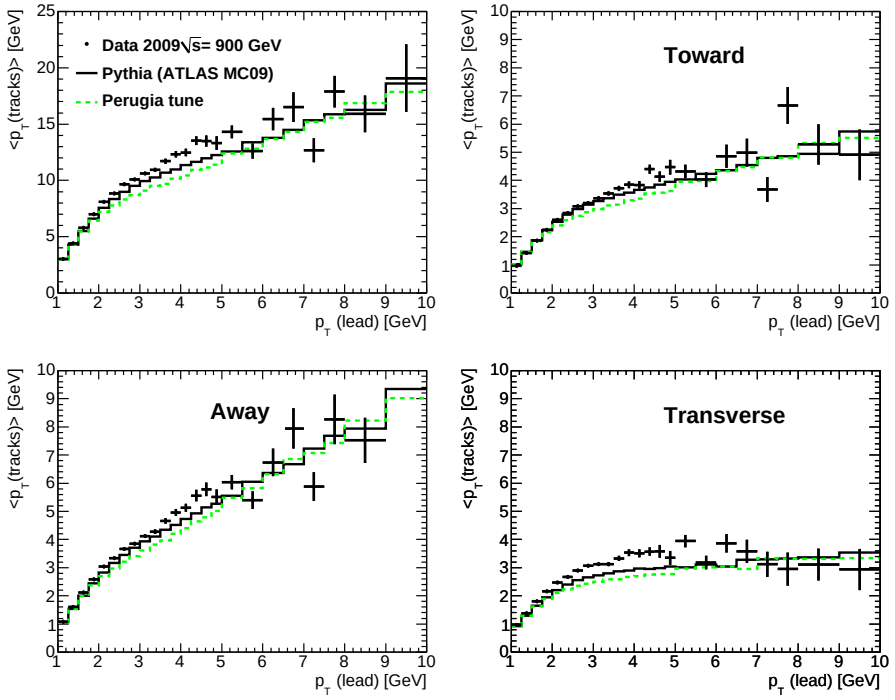

Figure 7: The average track transverse momentum per event calculated using tracks with $p_T > 0.5$ GeV as a function of $p_T(lead.)$ for the different regions defined in Fig. 1.

5 Data unfolding using bin-by-bin corrections

Due to the complexity of the measured variables, a bin-by-bin correction procedure is used. The correction factors

$$C = \frac{\mathscr{A}^{\text{gen}}}{\mathscr{A}^{\text{det}}},$$

are evaluated separately for each observable $\mathscr{A}$. In the above expression, $\mathscr{A}^{\text{gen}}$ is calculated at the generator-level of PYTHIA MC09 and $\mathscr{A}^{\text{det}}$ is that at the detector-level of this model after the event and the track selection. The corrected value for an observable is found by multiplying its measured value by the relevant correction factor.

The correction factors unfold the data to the hadron level and include a correction for event selection, efficiency corrections, purity corrections, bin-by-bin migrations, smearing of the distributions when the leading particle is misidentified (typically, this leads a smearing of particle densities for the $\delta\phi$ variable).

In case if $\mathscr{A}$ is a simple particle-counting observable, the bin-by-bin correction can be represented as a ratio of the purity to the reconstruction efficiency:

$$C = \rho/e,$$

where ρ is a purity calculated as the ratio:

$$\rho = \frac{N(reco\,\&\,gen)}{N(reco)},$$

where $N(reco\,\&\,gen)$ is the number of reconstructed tracks which originate from truth particles generated in same bin. $N(reco)$ is the number of events with reconstructed tracks counted in the same bin,

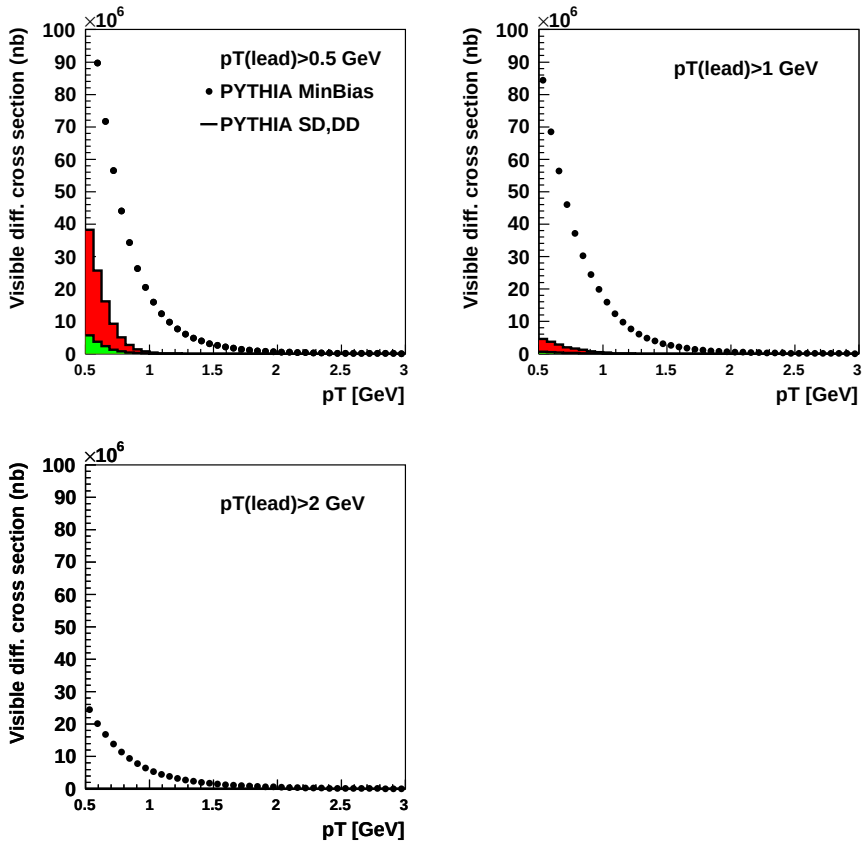

Figure 8: The visible differential cross section for tracks calculated using the PYTHIA model. The largest shaded area (in red) shows the contribution from single diffraction, while the green histogram (smallest contribution) shows the contribution from double diffractive.

irrespective of their origin. The efficiency is defined as usual:

$$e = \frac{N(reco\,\&\,gen)}{N(gen)},$$

where $N(gen)$ is the number of generated truth particles in the same bin where $N(reco\,\&\,gen)$ is reconstructed. While the efficiency is directly calculable and includes an event-selection and standard track-reconstruction efficiency. Both efficiency and purity have been estimated in [5]), the purity mainly reflects smearing effects due to miss-measurements of the leading tracks used to for the density calculations which are difficult to take into account. Thus the advantage of using the bin-by-bin correction is that it unfolds data in one step. [3]

Figure 9 shows the $\delta\phi$ density for charged particles at the truth level and at the detector level (tracks). The ratio of those (i.e. the bin-by-bin correction factors) are shown in the bottom figures. The correction factors are close to 1.2 for all observables and vary smoothly for any given $\mathscr{A}$. The source of the deep at $\delta\phi = 0$ is not yet fully understood.

Similarly, Figs. 10, 11 and 12 show the densities at the truth level (stable charged particles) and the detector level (for tracks). The corresponding bin-by-bin correction factors are also shown. The correction factors are close or below 20% for all distributions, they are smooth and vary only slowly with $p_T(lead.)$

6 Systematic uncertainties

The systematic uncertainties on the measured densities were determined by changing the selection cuts or the analysis procedure and repeating the analysis. The following systematic studies have been carried out, with a typical resulting uncertainty for the given in parentheses:

- A total uncertainty on track reconstruction efficiency was assumed to be 3% [5]. This uncertainty was included by scaling the bin-by-bin correction factors assuming that the purity factor is independent of the efficiency.

  This uncertainty is significantly larger than the event selection uncertainty due to the trigger selection of the minimum bias events and the vertex position [5].

- Since no correction on the diffractive contribution was applied, the measurement was performed in the regions of the $p_T(lead.)$ where the diffractive contribution is small (see Sect. 4).

  It was checked how the rejection of low-multiplicity events, which are mostly influenced by the diffractive contribution, affects the final measurements. For this, all events with the number of primary tracks below 3 were removed. ($< 1\%$.)

- the minimum transverse momentum for tracks was varied by 1% in the data to take into account a difference between data and Monte Carlo in resolution. ($\pm 1\%$.)

- track selection was tightened for data and Monte Carlo by reducing the $d0$ and $z0\sin(\theta)$ cuts to 1 mm, and by tightening the cuts on the SCT hits. ($< 2\%$.)

- The bin-by-bin corrections were estimated using a MC with an extra 10% material in front of the tracking system. The extra material decreases the efficiencies and thus increases the bin-by-bin correction factor. $(+2.5\%)$.

[3] It should be pointed out that the truth level used for the bin-by-bin correction does not contain diffractive events. This means that the bin-by-bin correction factor corrects the data for the presence of diffraction at the reconstructed level.

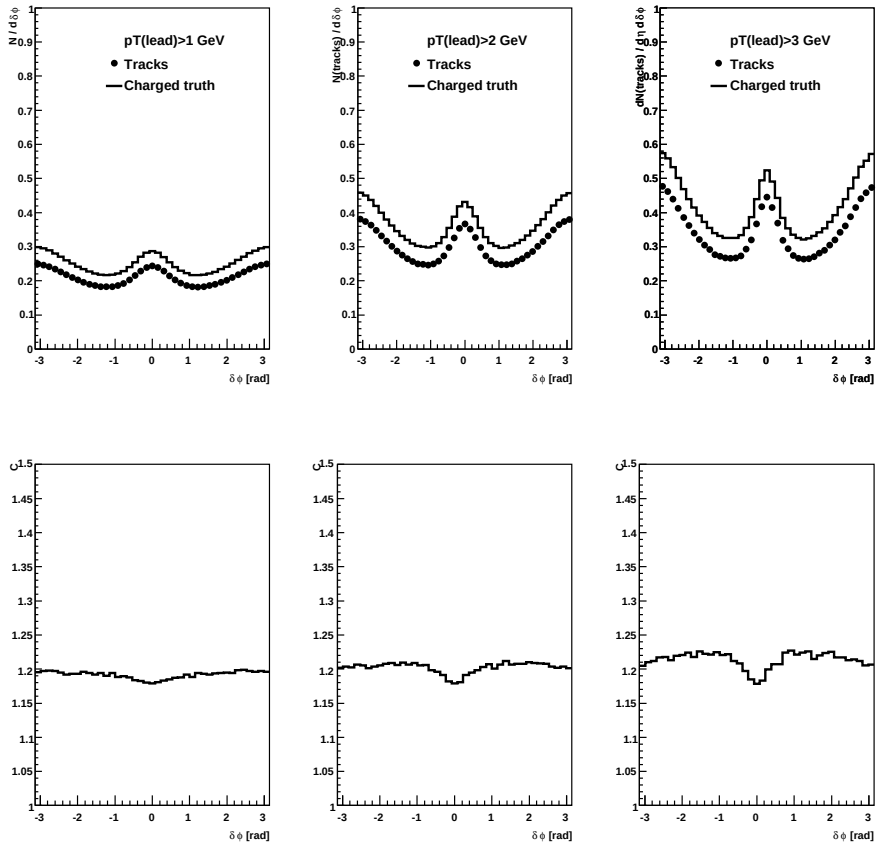

Figure 9: Densities for the truth level (charged stable particles) and the detector levels (tracks) as a function of $\delta\phi$. The bin-by-bin correction factors are shown at the bottom.

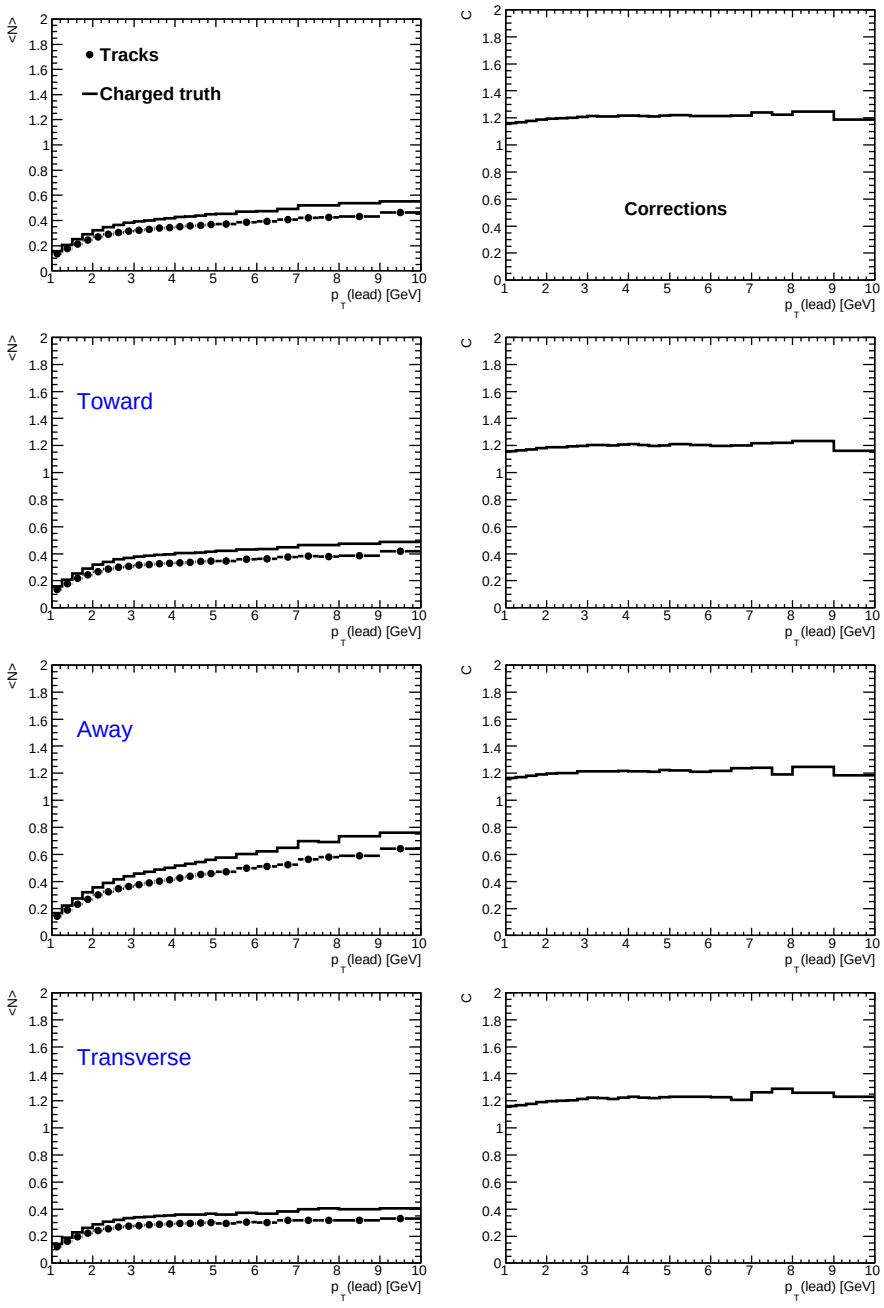

Figure 10: Densities for the truth level (charged stable particles) and the detector levels (tracks) as a function of $p_T(lead.)$. The bin-by-bin correction factors are shown on the left side.

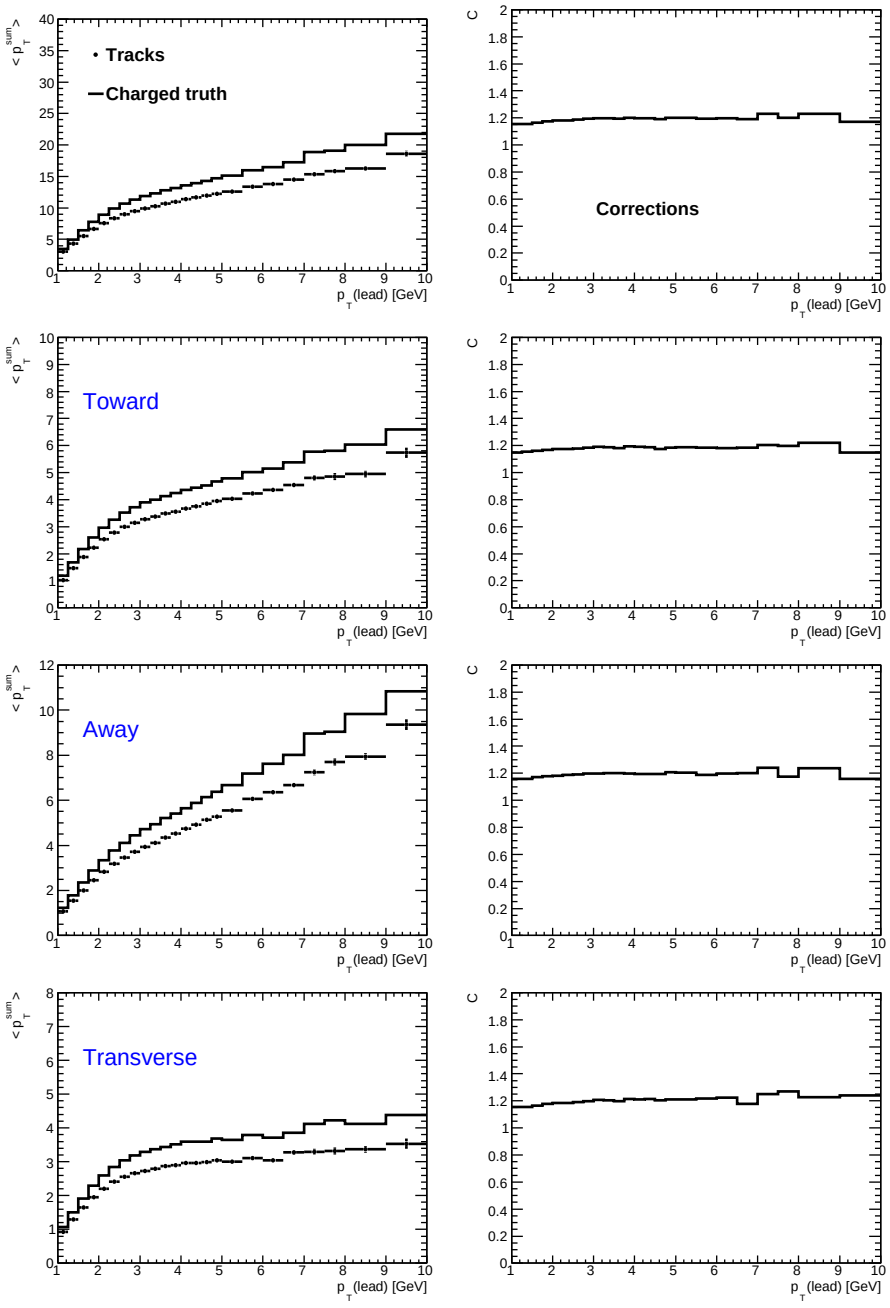

Figure 11: The average transverse momenta for the truth level (charged particles) and the detector levels (tracks)as a function of $p_T(lead.)$. The bin-by-bin correction factors are shown on the left side.

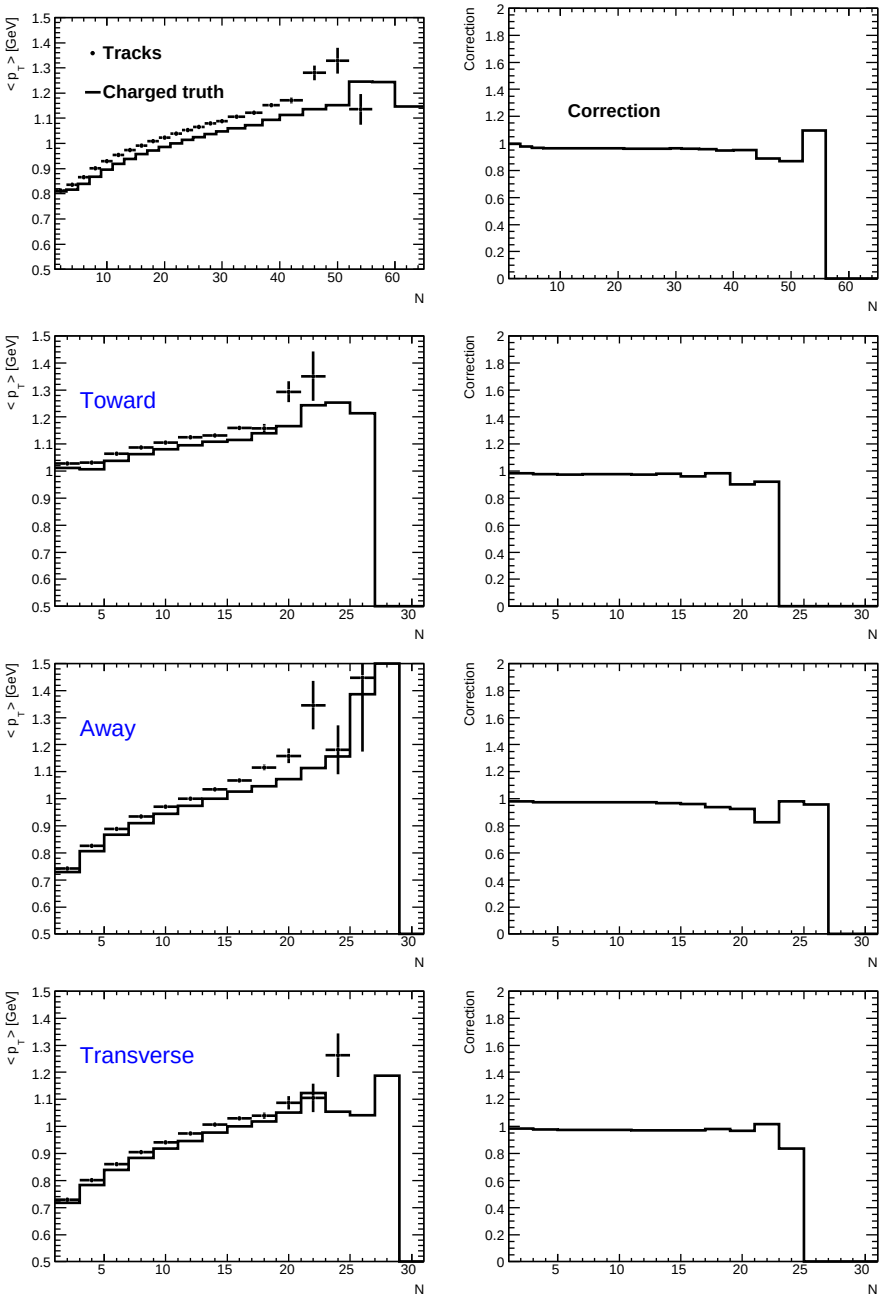

Figure 12: The average transverse momenta for the truth level and the detector levels (tracks) as a function of multiplicities. The correction factors are shown on the left side.

- Several Monte Carlo models with different alignments were used ($< 0.5\%$).

  This uncertainty and the uncertainty due to the 10% extra material discussed above are already included in the 3% uncertainty on the tracking efficiency which was used to scale the bin-by-bin correction factor (see the first bullet). This will result in a conservative estimate of the measurement uncertainty due to a double counting of uncertainties. However, this was necessary in order to verify the effect on the purity due to misidentification of the leading track.

- A model dependence on the bin-by-bin corrections was estimated using alternative PYTHIA tunes (Perugia0 and DW). ($< 1\%$.)

We did not include a possible systematic uncertainty from the PHOJET Monte Carlo model [3] since it significantly fails to describe the data at large $p_T(lead.)$ [5]. For the bin-by-bin corrections, an adequate description of detector-level distributions is required, thus the generator-level of PHOJET should be re-weighted before the extraction of the correction. The PHOJET MC has exactly the same fragmentation as for PYTHIA, thus we do not expect this model to be useful for estimation of the systematical uncertainties related different modeling of the fragmentation stage. The HERWIG Monte Carlo model [6, 7] is presently unavailable.

The overall systematic uncertainty was determined by adding the above uncertainties in quadrature.

7 Results

Figure 13 shows the density distribution of the charged particles corrected to the hadron level as a function of the distance in the azimuthal angle between the leading charged particle and other charged particles in an event. The distribution was unfolded using the bin-by-bin correction as shown in Fig. 9. The systematic uncertainties are almost 100% correlated. They are shown by the yellow band which also includes the statistical errors added in quadrature. The data are compared to the PYTHIA truth with the MC09c, Perugia0 and DW tunes. Although the general shapes of the Monte Carlo distributions are similar to that of the data, none of the three Monte Carlo tunes match the data precisely.

Figures 14 and 15 show the charged particle density and the average momenta of charged particles corrected to the hadron level as a function of $p_T(lead.)$. Figures 16 and 17 show the average momenta as a function of charged multiplicity, and the sum of the p_T. While there are small differences in most of the density and momentum distributions, it is important to note that this is not the case for the sum p_T. In this case the distribution is well described by all Monte Carlo predictions. Although this implies that the distribution is of little use in Monte Carlo tuning, it does imply that there is little model dependence in modeling the underlying event as is necessary for example in the determination of the jet energy scale.

8 Summary

In this note, density and the average transverse spectra are studied and compared with the PYTHIA predictions with different tunes. All predictions fail to describe well the density distribution as a function of the azimuthal angle between the leading charged particle and any other particle in an event. In particular, the Monte Carlo predictions underestimate the hadronic final state activity in the transverse region and whose disagreement increases with the p_T of the lead particle. Similarly, the lack of activity of the hadronic final state in the transverse region is seen in the density distribution and the average transverse momenta as a function of the $p_T(lead.)$. We note however, that all tunes describe well the sum p_T in the events, which indicates that there is little model dependence in the estimate of the global underlying event energy as is needed for the determination of the jet energy scale.

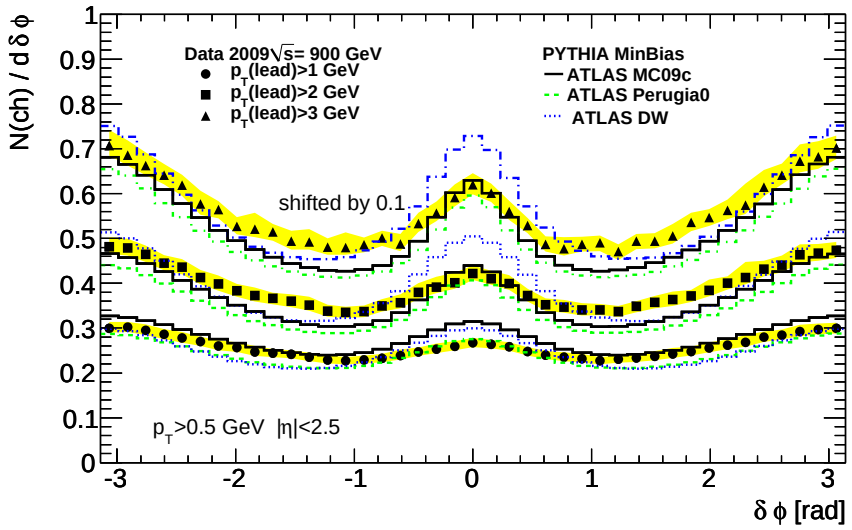

Figure 13: The density of charged particles as a function of the distance between the leading particle and other particles in an event. The yellow band shows the statistical and systematical uncertainties added in quadrature. The density for $p_T(lead.) > 3$ GeV is shifted by 0.1.

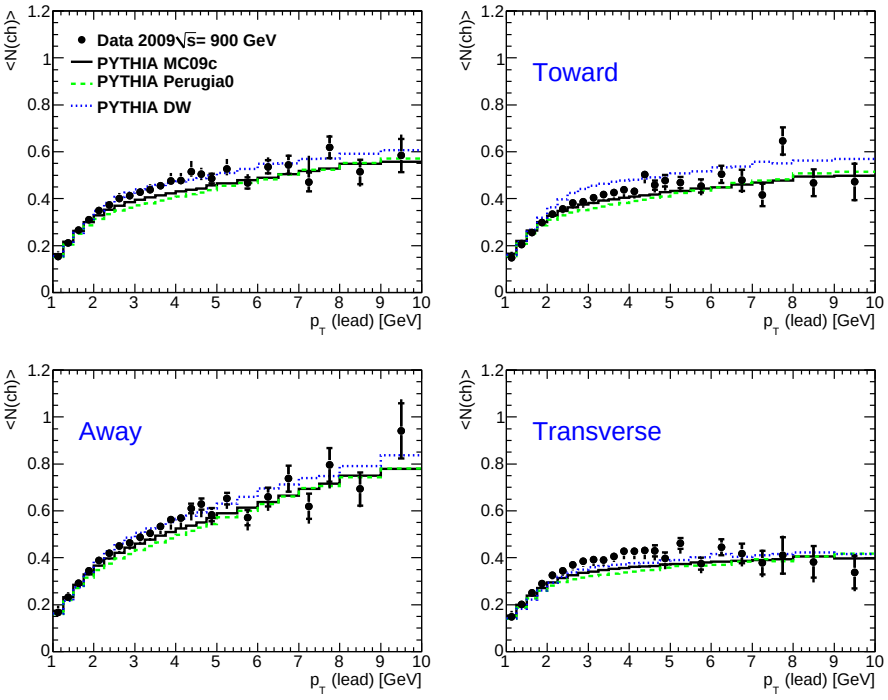

Figure 14: The average number of tracks per event in one unit interval in η and ϕ as a function of the $p_T(lead.)$ for different regions of the phase space indicated in Fig. 1

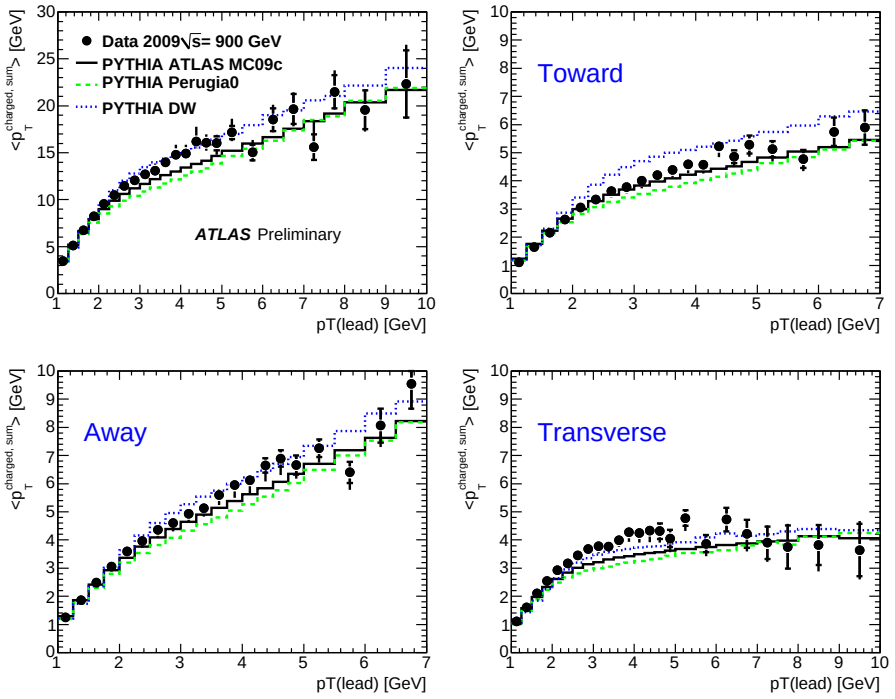

Figure 15: The average transverse momenta of charged particles as a function of the $p_T(lead.)$ for different regions of the phase space indicated in Fig. 1.

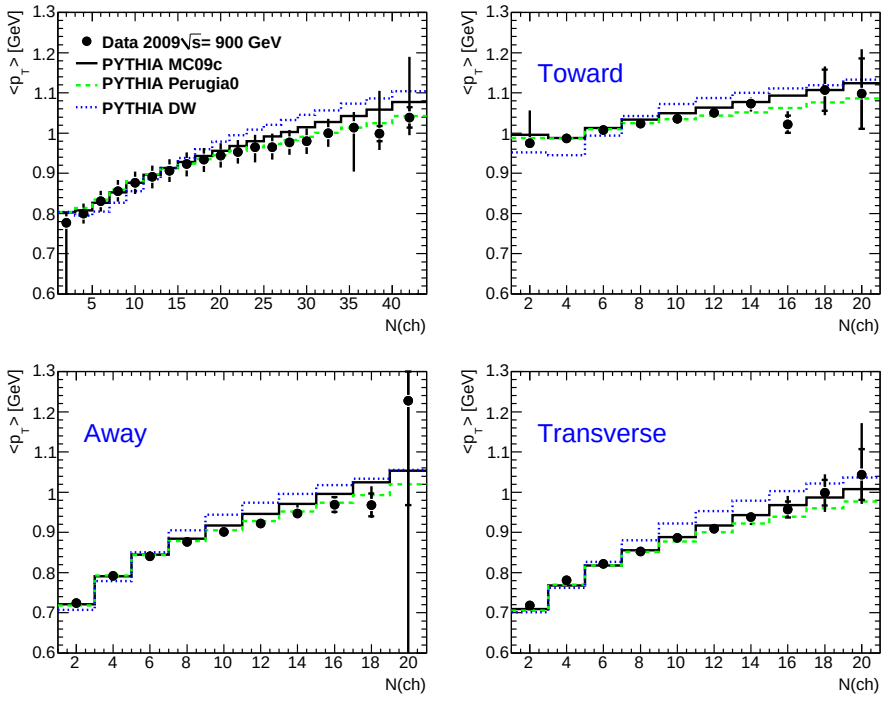

Figure 16: The average transverse momenta of charged particles as a function of charged-particle multiplicity for different regions of the phase space indicated in Fig. 1.

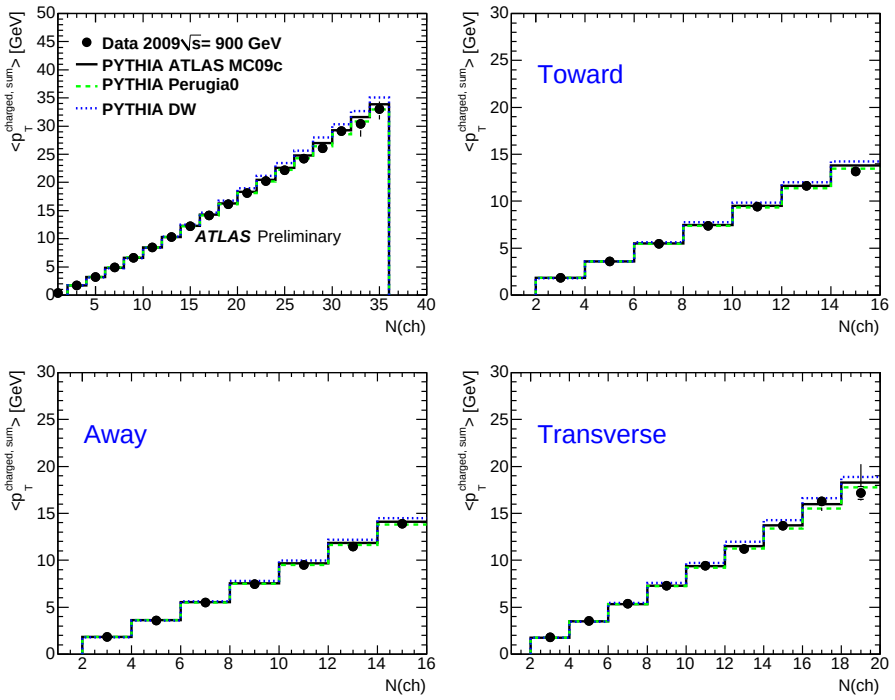

Figure 17: The sum of the transverse momenta of all charged particles as a function of charged-particle multiplicity for different regions of the phase space indicated in Fig. 1.

References

[1] The ATLAS Collaboration, *Charged-particle multiplicities in pp interactions at sqrt(s) = 900 GeV measured with the ATLAS detector at the LHC*, arXiv:1003.3124 (2010) .

[2] NtupleMaker, *NtupleMaker package (ANL-ASC)*, . `http://atlaswww.hep.anl.gov/asc/WebSVN/`.

[3] R. Engel and J. Ranft, *Color singlet exchange between jets and the PHOJET Monte Carlo*, Nucl. Phys. Proc. Suppl. **75A** (1999) 272–274.

[4] NtupleMakerTruth, *NtupleMakerTruth package (ANL-ASC)*, `http://atlaswww.hep.anl.gov/asc/WebSVN/`.

[5] The ATLAS Collaboration Collaboration, A. ATLAS, *Charged-particle multiplicities in pp interactions at sqrt(s)=900 GeV measured with the ATLAS detector at the LHC*, arXiv:1003.3124 (2010) .

[6] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **01** (2001) 010, `arXiv:hep-ph/0011363`.

[7] G. Corcella et al., *HERWIG 6.5 release note*, `arXiv:hep-ph/0210213`.